

WYKORZYSTANIE MODELU LINIOWEGO HOLTA DO PROGNOZOWANIA WYDATKÓW NA OCHRONĘ ZDROWIA

AGNIESZKA ROGÓŻ-DUDA

Streszczenie

W artykule zaprezentowano zastosowanie modelu Holta do prognozowania wydatków na ochronę zdrowia. Należy zwrócić szczególną uwagę na zbudowanie modelu prognostycznego i odpowiedni dobór parametrów dla tego modelu. Przedstawiona metoda daje możliwość analizy problemu przy zastosowaniu trzech różnych zestawów parametrów. Wartości powinny być dobrane poprzez minimalizację wartości średniego błędu kwadratowego prognoz wygasłych. Autorka dobrała parametry optymalne oraz w sposób przypadkowy. Celem prowadzonych badań było zaprezentowanie skutków niewłaściwego doboru parametrów modelu.

Słowa kluczowe: model liniowy Holta, parametry modelu, nakłady finansowe na ochronę zdrowia

Wprowadzenie

Badanie szeregów czasowych względem ochrony zdrowia ma dwa przeznaczenia. W pierwszym umożliwia zapoznanie się z informacjami dostępnymi dla ogółu ludności w Polsce odnośnie nakładów finansowania jakie ponosi Narodowy Fundusz Zdrowia na ochronę zdrowia. W drugim natomiast po zdefiniowaniu modelu prognostycznego oraz błędów odpowiednie instytucje medyczne byłyby w stanie zaplanować swoje wydatki na krótkie okresy rozliczeniowe.

Wygładzanie wykładnicze jest wykorzystywane do skorygowania szeregu czasowego prognozy. Korzyścią zastosowania modeli tego rodzaju jest ich charakter rekurencyjny, przy czym należy założyć, że prognozy uaktualniają się po dodaniu nowych danych [2, s. 285–300]. Jedną z metod jest model liniowy Holta, który wykorzystuje się do analizy danych historycznych [10, 12].

1. Liniowy model Holta

Model liniowy Holta, który polega na wygładzaniu badanego szeregu czasowego przy pomocy średniej ruchomej, w której występuje zarówno tendencja rozwojowa jak i wahania przypadkowe używany jest do prognozowania krótkookresowego dla m kroków w przód. Nie należy jednak wydłużać zakresu czasowego, ponieważ wiąże się to ze wzrostem niepewności [11, s. 24–26]. Dodatkowo wiadomo, że każda prognoza powoduje niedokładność, ale umożliwia również pokazanie celu lub kierunku do którego należy dążyć. Nie udziela jednak rozwiązania na podane problemy [15, s. 73].

Wartości prognozowania oznaczone są jako y_0, y_1 aż do y_{n-1} i związane są z parametrami α i β . Równania modelu przyjmują następujące postaci [14, s. 16–17]:

$$F_t = \alpha y_t + (1 - \alpha)(F_{t-1} + S_{t-1}) \quad (1)$$

$$S_t = \beta(F_t - F_{t-1}) + (1 - \beta)S_{t-1} \quad (2)$$

$$F_1 = y_1 \quad (3)$$

$$S_1 = y_1 - y_0 \quad (4)$$

gdzie:

t – kolejny okres czasu ($t = 2, \dots, n-1$),

F_t – wygładzona wartość analizowanego szeregu czasowego w okresie t ,

S_t – wygładzona wartość trendu w okresie t ,

α, β – parametry modelu.

Prognozy kolejnych wartości szeregu określa się za pomocą wzoru 5:

$$y_{n+k-1}^* = F_{n-1} + k \cdot S_{n-1}; k = 1, 2, \dots \quad (5)$$

Ponadto dla równania 5 należy założyć, że $n = t$ oraz $k = 1$ otrzymując:

$$y_t^* = F_{t-1} + S_{t-1} \quad (6)$$

Następnie prognozy porównywane są z danymi rzeczywistymi, a różnicą są błędy prognoz wygasłych z założonymi parametrami. Wybór parametrów α i β opiera się na minimalizacji błędów prognoz wygasłych. Powszechnie przyjmują one wartości od 0 do 1.

2. Metoda badań

Stworzenie odpowiedniej prognozy opierało się na zbudowaniu modelu prognostycznego, tak aby uzyskać wydatki ogółem na ochronę zdrowia na lata 2015–2018. Przeprowadzono analizę zgromadzonych danych historycznych za lata 2007–2014, a następnie zaobserwowano element główny związany z tendencją rozwojową oraz składową losową o charakterze wahań przypadkowych. Informacje uzyskane z okresu poprzedzającego prognozy uwzględniają nakłady finansowe ogółem (tj. wydatki bieżące i inwestycyjne). Wydatki bieżące rozumiane są jako koszty związane z indywidualną opieką zdrowotną, czyli na ochronę zdrowia. Zaliczyć do nich można: usługi lecznicze, pomocnicze i rehabilitacyjne, długotrwałą opiekę pielęgniacyjną oraz leki i produkty medyczne. Wydatki inwestycyjne to zbiór wydatków związany z ochroną zdrowia, a więc inwestycje, kształcenie personelu medycznego, badania i rozwój w ochronie zdrowia oraz administracja świadczeń pieniężnych [3–9]. Stosunek wydatków jest zależny od realizowanych funkcji. Przegląd wyników Narodowego Funduszu Zdrowia na lata 2003–2008 dowodzi, że nastąpił regularny wzrost wydatków ogółem, jednak z utrzymaniem struktury na wydatki publiczne i prywatne. Badanie szeregów czasowych wydatków na ochronę zdrowia w Polsce przedstawia tabela 1.

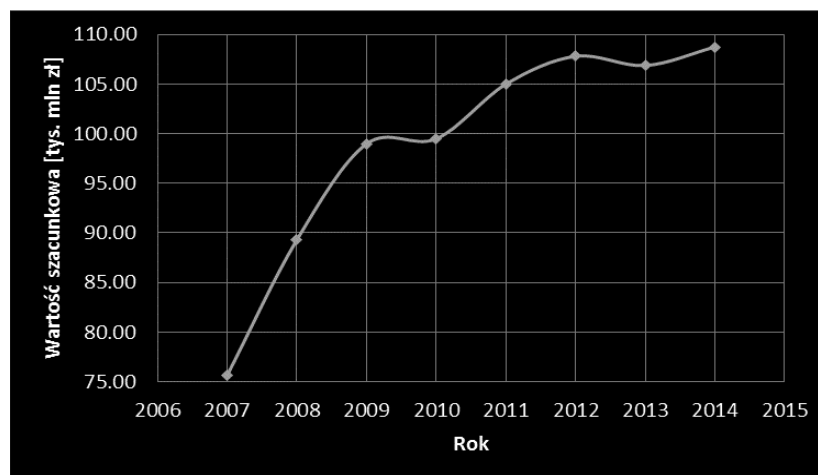
Tabela 1. Wydatki ogółem na ochronę zdrowia w Polsce w latach 2007–2014

Rok	Wydatki ogółem na ochronę zdrowia (mln zł)
2007	75 665,00
2008	89 307,00
2009	98 976,00
2010	99 485,00
2011	104 997,00
2012	107 802,00
2013	106 869,40
2014	108 737,20

Źródło: [3–9].

Po przeanalizowaniu danych historycznych zauważono, że różnica pomiędzy wydatkami na ochronę zdrowia w latach 2007–2008 była największa. Wynosiła ona aż 15,28% wzrostu, co daje 13 642 mln zł. Natomiast w okresie 2012–2013 zauważyć można spadek nakładów finansowych. Średni wzrost w badanych latach wynosił 4724,60 mln zł. Od roku 2011 jednak wzrost/spadek wydatków utrzymywany był na poziomie $-0,87\%$ do $+2,60\%$, odpowiednio w: 2011 roku $+2,60\%$, 2012 roku $-0,87\%$ oraz 2013 $+1,72\%$.

Analizowany szereg czasowy przedstawiono na rysunku 1.



Rysunek 1. Dane rzeczywiste w szeregu czasowym

Źródło: opracowanie własne na podstawie [3–9].

Przeprowadzając wizualną ocenę danych rzeczywistych można zauważyć, że wykorzystanie modelu Holta do prognozowania wydatków na ochronę zdrowia było trafnie dobrane do przebiegu analizowanego szeregu czasowego. Warto zauważyć, że model ten jest używany najczęściej spośród

metod wykładniczego do budowania prognoz krótkookresowych. Za względu na składowe widoczny jest wzrostowy trend liniowy oraz wahania przypadkowe w 2010 i 2013 roku.

Najważniejszą częścią po zbudowaniu modelu jest oszacowanie prognozy i jej przegląd. Do oceny prognozy warto zastosować mierniki charakterystyczne dla obszaru zastosowań oraz wskaźniki uniwersalne. W niniejszym artykule wykorzystano następujące miary opisu statystycznego [1, s. 240, 13, s. 36]:

- MAE (ang. *mean absolute error*) – średni błąd bezwzględny,
- MAPE (ang. *mean absolute percentage error*) – średni bezwzględny błąd procentowy,
- RMSE (ang. *root mean square error*) – pierwiastek błędu średniokwadratowego),
- V_{RMSE} (ang. *relative prediction error ex-post*) – względny błąd predykcji ex post.

3. Wyniki badań

Składowe do prognozowania przeanalizowano za pomocą modelu liniowego Holta, gdzie parametry metody są uzależnione od założonych miar jakości. Wartości te zostały dobrane z dokładnością do 0,0001. Określono prognozy na cztery przyszłe okresy, przy czym parametry optymalne zostały wyznaczone za pomocą dodatku Solver pakietu MS Excel. Wartości te zostały dobrane poprzez minimalizację wartości średniego błędu kwadratowego prognoz wygasłych. Użytkowano następujące wartości parametrów: $\alpha = 0,8981$ oraz $\beta = 0,5106$.

Wygenerowane prognozy porównano do wartości par dwóch parametrów modelu: $\alpha = 0,8$ i $\beta = 0,2$ oraz $\alpha = 0,2$ i $\beta = 0,8$ w celu wskazania różnic, jakie mogą występować przy nieodpowiednim dobraniu parametrów do modelu.

Wartości obliczone dla optymalnych parametrów modelu zaprezentowano w tabeli 2.

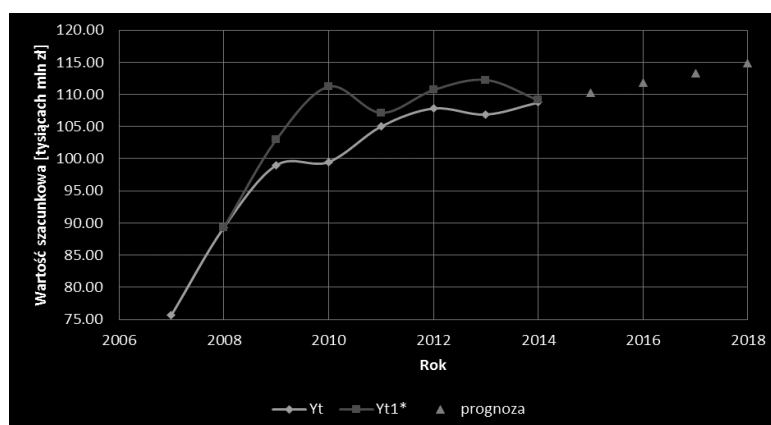
Tabela 2. Wyniki obliczeń

t	Y_t	F_t	S_t	Y_{t1}^*	Błąd prognozy wygasłej
2007	75 665,00	75 665,00	13 642,00		
2008	89 307,00	89 307,00	13 642,00	89 307,00	0,00
2009	98 976,00	99 380,92	11 820,23	102 949,00	-3 973,00
2010	99 485,00	100 679,07	6 447,92	111 201,14	-11 716,14
2011	104 997,00	105 214,08	5 471,24	107 127,00	-2 130,00
2012	107 802,00	108 095,86	4 149,12	110 685,32	-2 883,32
2013	106 869,40	107 417,26	1 684,21	112 244,98	-5 375,58
2014	108 737,20	108 774,33	1 517,18	109 101,47	-364,27
2015				110 291,50	
2016				111 808,68	
2017				113 325,86	
2018				114 843,03	

Źródło: opracowanie własne.

Wartości rzeczywiste i prognozy różnią się między sobą, co odzwierciedla błąd prognozy wygasłej. Waha się on od 11 716,14 mln zł do 2 130,00 mln zł, co świadczy o różnicy pomiędzy daną rzeczywistą a prognozą. Im większa rozbieżność pomiędzy nimi tym większy błąd prognozy wygasłej.

Graficzna interpretacja szeregu czasowego oraz prognozy zilustrowana została na rysunku 2. W tabeli 3 zaprezentowano wyniki obliczeń z porównania trzech modeli o różnych wartościach parametrów α i β .



Rysunek 2. Prognoza dla $\alpha = 0,8981$ oraz $\beta = 0,5106$.

Źródło: opracowanie własne.

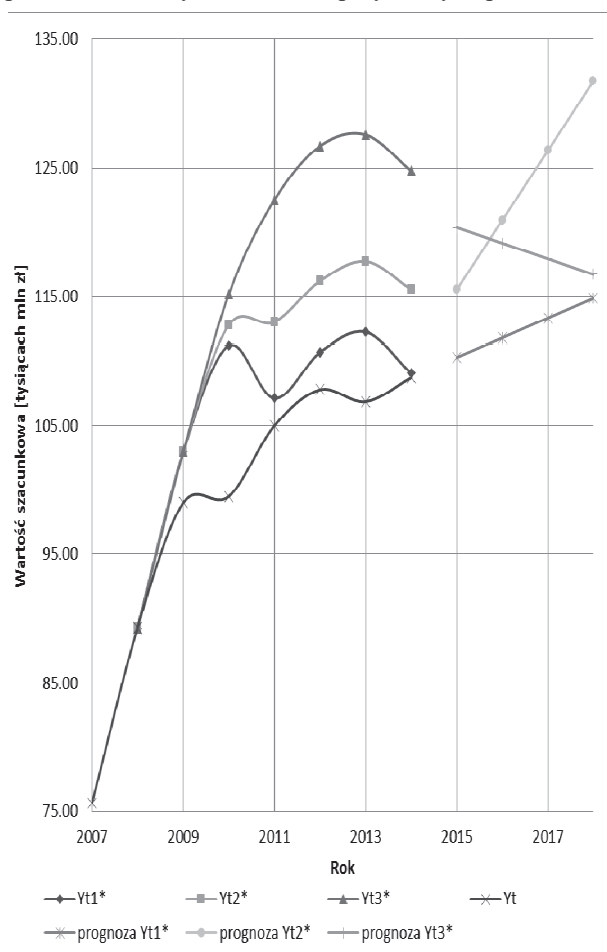
Tabela 3. Wyniki obliczeń – porównanie analizowanych modeli

t	Dane rzeczywiste	Model 1			Model 2			Model 3		
	Y_t	F_t	S_t	Y_{t1}^*	F_t	S_t	Y_{t2}^*	F_t	S_t	Y_{t3}^*
2007	75665,00	75665,00	13642,00		75665,00	13642,00		75665,00	13642,00	
2008	89307,00	89307,00	13642,00	89307,00	89307,00	13642,00	89307,00	89307,00	13642,00	89307,00
2009	98976,00	99380,92	11820,23	102949,00	99770,60	13006,32	102949,00	102154,40	13006,32	102949,00
2010	99485,00	100679,07	6447,92	111201,14	102143,38	10879,61	112776,92	112025,58	10498,20	115160,72
2011	104997,00	105214,08	5471,24	107127,00	106602,20	9595,45	113023,00	119018,42	7693,92	122523,78
2012	107802,00	108095,86	4149,12	110685,32	109481,13	8252,15	116197,65	122930,28	4668,26	126712,34
2013	106869,40	107417,26	1684,21	112244,98	109042,18	6513,93	117733,28	123452,71	1351,60	127598,54
2014	108737,20	108774,33	1517,18	109101,47	110100,98	5422,90	115556,10	121590,89	-1219,14	124804,31
2015				110291,50			115523,88			120371,76
2016				111808,68			120946,79			119152,62
2017				113325,86			126369,69			117933,48
2018				114843,03			131792,59			116714,35

Źródło: opracowanie własne.

Parametry modelu 1 zostały dobrane jako optymalne, natomiast w przypadku modelu 2 i 3 autorka nie wyznaczała ich, lecz dobrała w sposób przypadkowy ich wartości. Model 1 charakteryzuje się najlepszymi dopasowaniami do danych rzeczywistych. Model Holta w tym przypadku został dobrany prawidłowo, ponieważ najdokładniej wyznaczył prognozy wygasłe.

Na rysunku 3 zaprezentowano wyniki obliczeń przy różnych parametrach α i β .



Rysunek 3. Prognoza dla różnych parametrów α i β

Źródło: opracowanie własne.

Na rysunku 3 przedstawiono poglądowo jak wygląda interpretacja tabeli 3. Widoczne jest, że prognozy Y_{t2}^* oraz Y_{t3}^* są różne od siebie i odbiegają znacząco od prognozy Y_{t1}^* , która posiada optymalny dobór parametrów α i β . Po zdefiniowaniu prognoz dla trzech różnych parametrów α i β wyznaczono ich ocenę dokładności i trafności. Założono, że prognoza nie może zostać obciążona błędem większym niż 5%. Określone błędy porównano w tabeli 4.

Tabela 4. Błędy prognoz

Model	1	2	3
Parametry modelu	$\alpha = 0,8951$ i $\beta = 0,5106$	$\alpha = 0,8$ i $\beta = 0,2$	$\alpha = 0,2$ i $\beta = 0,8$
MAE [mln zł]	3 777,47	7 338,48	13 268,87
MAPE [%]	3,69	7,20	12,60
RMSE [mln zł]	5 277,08	8 382,68	15 185,29
V _{RMSE} [%]	4,97	7,65	13,14

Źródło: opracowanie własne.

Prognoza wydatków na ochronę zdrowia w 2014 osiągnęła wartość 109 101,47 mln zł i w porównaniu do danych rzeczywistych odbiegała ona tylko o 0,34%. Największą różnicę zaobserwowano w 2010 roku, przy czym wynosiła ona 11,78% od stanu faktycznych nakładów finansowych.

4. Podsumowanie

Przy doborze optymalnych parametrów modelu liniowego Holta ($\alpha = 0,8951$ i $\beta = 0,5106$) określono prognozę na lata 2015–2018, czyli na najbliższe 4 okresy rozliczeniowe. Błąd MAE wynoszący 3 777,47 wskazuje o ile średnio zrealizowane wydatki prognozowania mogą się różnić od prognoz. Przedstawiono przewidywane realizacje i oszacowano błąd MAPE wynoszący 3,69%. Średni względny błąd 7 prognoz wygasłych dla analizowanego okresu wskazuje je na dopuszczalne, przy czym założono, że nie może ona być obciążona błędem większym niż 5%. W modelu o optymalnych parametrach występuje najmniejszy stopień obciążenia błędem (4,97%) w porównaniu do modelu o parametrach $\alpha = 0,8$ i $\beta = 0,2$ (7,65%) oraz $\alpha = 0,2$ i $\beta = 0,8$ (13,14%). Jednak w modelu 2 różnica pomiędzy MAE a RMSE jest mniejsza niż w pozostałych modelach i wynosi 1 044,20 mln zł. Świadczy to o mniejszej różnicy pomiędzy prognozą a wartością prognozowaną. W przypadku modelu 1 równa się ona 1 499,60 mln zł, a w modelu trzecim 1 916,42 mln zł, co wskazuje na znaczne różnice pomiędzy tymi modelami. Pomimo tego wartość α najlepiej została dobrana dla pierwszego przypadku. Całość świadczy o tym, że ważny jest dobór parametrów tak by współczynnik RMSE był jak najmniejszy. Warto też zwrócić uwagę na różnicę MAE i RMSE by była najlepiej znikoma, ponieważ zbyt duża rozbieżność powoduje występowanie dużych wartości błędów.

Bibliografia

- [1] Chodakowska E. i in.: *Zastosowanie modeli klasy GARCH do prognozowania cen energii elektrycznej na Towarowej Gieldzie Energii SA*. [w:] Kiełtyka L., Nazarko J. (red.) *Technologie informatyczne i prognozowanie w zarządzaniu: wybrane zagadnienia*. Wydawnictwo Politechniki Białostockiej, Białystok 2005.
- [2] Gelper S., Fried R., Croux Ch.: *Robust Forecasting with Exponential and Holt-Winters Smoothing*. Journal of Forecasting, Vol. 29, 2010, s. 285–300.
- [3] <http://stat.gov.pl/obszary-tematyczne/zdrowie/zdrowie/narodowy-rachunek-zdrowia-za-2008-rok,4,1.html> [dostęp: 10 czerwca 2017 r.].
- [4] <http://stat.gov.pl/obszary-tematyczne/zdrowie/zdrowie/narodowy-rachunek-zdrowia-za-2009-rok,4,2.html> [dostęp: 10 czerwca 2017 r.].
- [5] <http://stat.gov.pl/obszary-tematyczne/zdrowie/zdrowie/narodowy-rachunek-zdrowia-za-20>

- 10-rok,4,3.html [dostęp: 10 czerwca 2017 r.].
- [6] <http://stat.gov.pl/obszary-tematyczne/zdrowie/zdrowie/narodowy-rachunek-zdrowia-za-2011-rok,4,4.html> [dostęp: 10 czerwca 2017 r.].
- [7] <http://stat.gov.pl/obszary-tematyczne/zdrowie/zdrowie/narodowy-rachunek-zdrowia-za-2012-rok,4,5.html> [dostęp: 10 czerwca 2017 r.].
- [8] <http://stat.gov.pl/obszary-tematyczne/zdrowie/zdrowie/zdrowie-i-ochrona-zdrowia-w-2013-r-,1,4.html> [dostęp: 10 czerwca 2017 r.].
- [9] <http://stat.gov.pl/obszary-tematyczne/zdrowie/zdrowie/narodowy-rachunek-zdrowia-2014,4,6.html> [dostęp: 10 czerwca 2017 r.].
- [10] Halicka K., Wieńkowski C.: *Wykorzystanie metod wygładzania wykładniczego do prognozowania kursu sprzedaży EUR*. *Ekonomia i Zarządzanie*, nr 2, 2013, s. 70–80.
- [11] Nazarko J. (red.): *Prognozowanie w zarządzaniu przedsiębiorstwem. Cz. 1 – Wprowadzenie do metodyki prognozowania*. Wydawnictwo Politechniki Białostockiej, Białystok 2004.
- [12] Nazarko J. (red.): *Prognozowanie w zarządzaniu przedsiębiorstwem. Cz. 2 – Prognozowanie na podstawie szeregów czasowych*. Wydawnictwo Politechniki Białostockiej, Białystok 2004.
- [13] Nazarko J. (red.): *Prognozowanie w zarządzaniu przedsiębiorstwem. Cz. 3 – Prognozowanie na podstawie modeli adaptacyjnych*. Wydawnictwo Politechniki Białostockiej, Białystok 2005.
- [14] Szostak R.: Uogólniony model Holta na przykładzie prognozowania liczby pasażerów w transporcie lotniczym w o Polsce, *Ekonometria*, nr 2, 2012, s. 16–26.
- [15] Zeliaś A.: *Teoria prognozy*. Polskie Wydawnictwo Ekonomiczne, Warszawa 1997.

**USE OF HOLT'S LINEAR MODEL FOR FORECASTING SPENDING
ON HEALTH CARE**

Summary

This article presents the use of the Holt model for forecasting health care expenditure. Pay special attention to the construction of the forecasting model and the appropriate selection of parameters for the model. The presented method gives you the opportunity to analyse the problem using three different sets of parameters. Values should be selected by minimizing the average square error value of the expired forecast. The author selected the optimal parameters and in random way. The purpose of this was to present what happens to the prognostic model after improper selection of parameters.

Keywords: Holt linear model, model parameters, financial outlay for health care

Agnieszka Rogóż-Duda
AGH Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie
Ul. Gramatyka 10, 30-067 Kraków
e-mail: agnieszka.rogocz@gmail.com